speech level.

In this apparatus, as the noise level in the environment in which the speaker is located changes between utterances, so his speech level is likely to rise and fall in accordance with the Lombard Effect, and the apparatus predicts the likely speech level. We have found that the likely speech level can be predicted with reasonable accuracy by measuring the noise immediately adjacent to an utterance; measuring the level of a steady noise is quite simple and can be carried out with just a short sample of the noise. The apparatus preferably also uses a measure of the speech level and the corresponding noise level relating to a previous or standardised utterance.

The ambient acoustic noise level could be measured before, after or even during utterance of a word or phrase, and it is preferred for the measurement to be made close in time to the utterance to reduce the possibility of the prediction of the likely speech level being inaccurate due to a significant shift in noise level between measurement and the actual utterance.

It is preferred for the measuring means to measure the ambient acoustic noise level immediately before the utterance, the estimate of speech level being determined before or as the utterance is made rather than thereafter. Alternatively the measurement may be after the utterance.

The apparatus preferably includes means operable to define, for each utterance, an utterance period comprising a first time period for measuring said acoustic noise level

and a second time period during which said utterance is made.

Thus in a preferred embodiment, the apparatus includes a user input device (such as e.g. a switch) and a timer and control means for defining said first noise measuring period, and said second speech measuring and/or recording period, the end of said first period being indicated to said user.

In a particularly preferred aspect, said apparatus is responsive to a succession of one or more utterances by a speaker and said measuring means measures the ambient noise level prevailing at each of said utterances to provide a series of noise measurements and said apparatus includes means for measuring the speech level of an utterance, and said processing means uses at least two of said noise measurements, together with the measurement of the speech level of the immediately previous utterance, to produce the prediction of the speech level of the most recent utterance.

In one example, where the noise is measured immediately before an utterance, the processing apparatus means predicts the speech level S_1^* of an utterance (1) on the basis of the following expression:

$$S_1^* = S_o + f(N_o - N_1)$$

where S_o is the speech level of the immediately previous utterance; N_1, N_o are the noise levels prevailing immediately before the utterance whose speech level is to be estimated, and immediately before the next previous utterance respectively, and $f(x)$ is a function relating changes in the

noise level in which the speaker is situated to the speaker's speech level.

The function is preferably monotonic increasing, and in a simple case is a multiplying factor less than 1. The multiplying factor may typically be a positive value in the range of from 0 to 0.6, and in one example is 0.32.

Alternatively the function may be a more complex function of the noise level difference. Likewise, the function may be modified to take account of more than just two noise level measurements; thus information relating to the speech levels of several previous utterances, together with the associated noise levels may be aggregated to predict the speech level of the next utterance.

In another aspect, this invention provides speech recognition or processing apparatus including predicting apparatus as set out above for use in adjusting the gain of the speech signal prior to recognition processing.

In yet another aspect, this invention provides a method for predicting the speech level of a speaker exposed to an environment containing a variable level of ambient acoustic noise, said method comprising the steps of:-

measuring said ambient acoustic noise level, and

processing said measured acoustic noise level to produce a prediction of the likely speech level.

In a further aspect, this invention provides a method for controlling the gain in a speech recognition or processing system, which comprises controlling the gain of the speech signal in accordance with a prediction of the

speech level obtained by the above method.

Whilst the invention has been described above, it extends to any inventive combination of the features set out above or in the following descriptions.

The invention may be performed in various ways, and an embodiment thereof will now be described by way of example only, reference being made to the accompanying drawing in which:-

Figure 1 is a block diagram of a speech recogniser incorporating speech level prediction in accordance with the invention.

The illustrated embodiment implements a system which applies knowledge of variation in the ambient acoustic noise level and its likely effect on the speech level to predict the speech level in the next utterance to be recognised by a speech recogniser. It is assumed that the variation in noise level over the duration of a single utterance is small compared with the variations occurring between utterances, and also that the noise has sufficient short-term stationarity that its level can be measured from a brief sample.

Referring to Figure 1, the speech recognition system comprises a microphone 10 whose output is subjected to voice processing at 12 before analogue to digital conversion at 14. The digital signal passes via a digital gain device 16 to a processor 18 which incorporates a recogniser 20 and a speech level estimator 22. The speech recogniser may be of any suitable type and examples of suitable recognisers will

be well known to those skilled in the art. The processor 18 also receives an input from a switch 24 acting as a user input device, and can issue warning tones to the user through a sounder 26.

The system illustrated is intended for use in a noisy environment whose noise level varies. In use, the user alerts the system when he wants to make an utterance to be recognised, by closing the switch 24. The processor then defines an utterance frame, comprising a first short time period, during which the ambient noise is sampled, followed by issuing a tone on the sounder 26, which indicates to the user that he may speak, followed by a second period during which the speech signal is sampled and sent to the recogniser 20. The second period is longer than the first period and sufficiently long to contain the longest utterance to be recognised. There are a number of ways of delimiting the second period other than providing a period of set duration. For example the length of the period may be user designated, e.g. by the user keeping the button pressed or pressing the button again. Alternatively, the processor may listen for a period of silence, or it may infer the end of a command based on an analysis of the grammar of the utterance. In addition, instead of using a switch, the start of the utterance frame may be marked by the user uttering a codeword.

Since it is known that speech levels vary with noise level, it is possible to predict a change in the speech level in an utterance from a change in the noise level. The

speech and noise levels, S_0 and N_0, (in dB units) are measured by the processor in one noise condition. The new noise level, N_1, in the first period of the next utterance, just before the start of an utterance to be recognised, is also measured by the processor. The difference in the two noise levels, N_0-N_1, is then determined and used by the processor, together with knowledge of the speech level, S_0 of the previous utterance, to predict the speech level, S_1, of the new utterance. We can write $S^*_1 = S_0 + f(N_0-N_1)$, where S^*_1 is a prediction estimate of S_1 and $f(x)$ is the function relating changes in the noise level in the speaker's ears to the speaker's speech level. In the simplest arrangement, the function is a multiplying factor less than 1, but it can also be a more complex function of the noise level difference. In practice we have determined empirically that the speech level good results are achieved in one application by using a multiplying factor of typically 0.3 although positive values between 0 and 0.6 should all provide some improvement. It may be assumed to be the same for all speakers or may be estimated separately for each speaker.

Since the measurements of the reference speech and noise levels, S_0 and N_0, respectively, are subject to measurement errors, it may be preferred to aggregate the information contributing to the prediction of S_1 from several previous utterances and noise estimates. The computation of S^*_1 described in the previous paragraph can be replaced by an average over several previous utterances. This may be a

simple average or it may be a weighted average, the weights possibly depending on factors such as the time difference between the various reference utterances and S_1 and on the relative durations of the various reference utterances. For example the computation may take account of any time effects. For example it may be found that, when exposed to a particular level of ambient noise that the speaker's speech level rises over an initial period and then decreases, in a temporal filtering effect.

Having determined an estimate of the speech level of the new utterance, the processor controls the gain of the signal accordingly. The gain may be adjusted at various points; it may be adjusted whilst the signal is still in the analogue domain or it may be achieved by digital scaling as shown by the digital gain device 16. A further alternative is to manipulate the fast fourier transform (FFT) values in the speech recogniser. If a cepstrum is computed, the signal may be scaled by adding an appropriate constant to the C_o coefficient. In a further arrangement, the system may compensate for increases or decreases in the speech level by adjusting the effective speech levels that the models in the recogniser represent.

The gain may take into account factors other than simply the level of the background noise; for example it could also take account of its spectral structure.

The output of the recogniser may be used in any convenient form. For example it could be used to enable a person to issue spoken commands to equipment.

Claims

1. Apparatus for predicting the speech level in an utterance of a speaker exposed to an environment containing a variable level of ambient acoustic noise, the apparatus comprising means for measuring said ambient acoustic noise level, and processing means for using said measured acoustic noise level to predict the likely speech level in said utterance.

2. Apparatus according to Claim 1, wherein said measuring means measures the ambient acoustic noise level immediately adjacent to said utterance.

3. Apparatus according to Claim 2, including means for activating said measuring means before the utterance.

4. Apparatus according to an preceding Claim which includes means operable to define, for each utterance, an utterance period comprising a first time period for measuring said acoustic noise level and a second time period during which said utterance is made.

5. Apparatus according to Claim 4, which includes a user input device, a timer, control means for defining said first period, and said second period, and means for indicating to a user the end of said first period.

6. Apparatus according to Claim 5, wherein said apparatus is responsive to a succession of one or more utterances by a speaker, and said measuring means is operable to measure the ambient noise level prevailing at each of said utterances to provide a series of noise values, and said

apparatus includes means for measuring the speech level of an utterance, and said processing means uses at least two of said noise values, together with a value representative of the speech level of the immediately previous utterance, to predict the likely speech level of the next utterance.

7. Apparatus according to Claim 6, wherein said measuring means is adapted to measure the ambient acoustic noise level before an utterance, and the processing means estimates the speech level S_1^* of an utterance (1) on the basis of the following expression:

$$S_1^* = S_o + f(N_o - N_1)$$

where

S_o is the speech level of the immediately previous utterance;

N_1, N_o are the noise levels prevailing immediately before the utterance whose speech level is to be estimated, and immediately before the next previous utterance respectively, and

$f(x)$ is a function relating changes in the noise level in which the speaker is situated to the speaker's speech level.

8. Apparatus according to Claim 7, wherein said processing means predicts the speech level S_1^* on the basis of the following expression:

$$S_1^* = S_o + k(N_o - N_1)$$

where k is a constant, $k>1$.

9. Apparatus according to Claim 8, wherein k lies in the range of from 0 to 0.6.

10. A speech recognition apparatus for use in an environment containing ambient acoustic noise, said apparatus including speech recogniser means for receiving and processing data representative of a speech utterance to be recognised to output data representative of or dependent on the lexical content of said utterance, said apparatus including level adjusting means for adjusting the level of the speech utterance, said apparatus further including means for measuring said ambient acoustic noise level before or during said utterance, processing means for using said measured acoustic noise level to predict the likely level of the speech utterance, and means for adjusting said level controlling means in accordance with said prediction of the likely level of the speech utterance.

11. A method for predicting the speech level of an utterance of a speaker exposed to an environment containing a variable level of ambient acoustic noise, said method comprising the steps of:-

measuring said ambient acoustic noise level, and

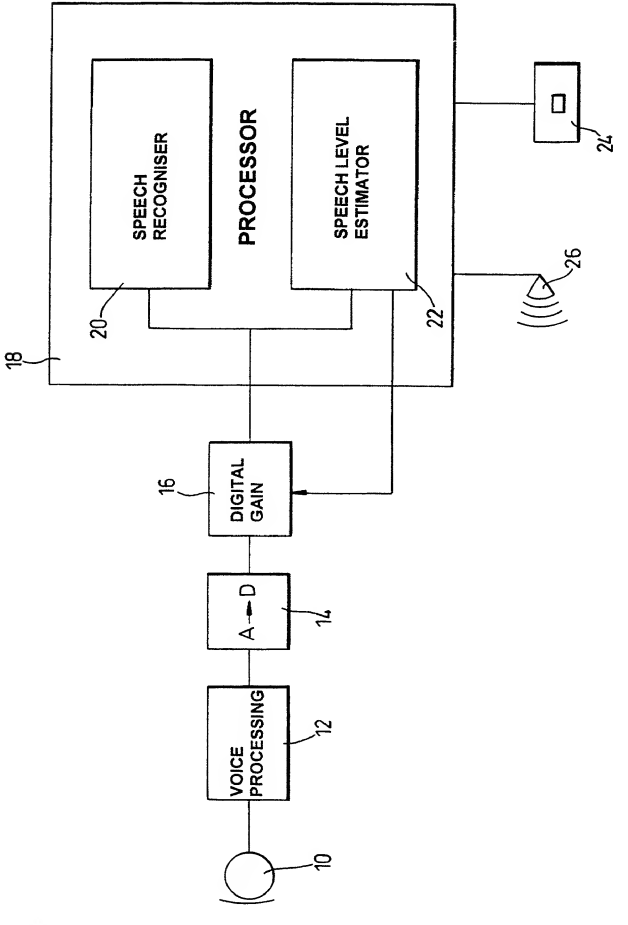
processing said measured acoustic noise level to predict the likely speech level of said utterance.

12 A method according to Claim 11, wherein said ambient acoustic noise level is measured before said utterance.

13. A method according to Claim 11, wherein a plurality of measurements of said acoustic noise level is taken and used with one or measurements of the speech levels corresponding to said measurements of acoustic noise level to predict the likely speech level of the utterance.

14. A method for controlling the gain in a speech recognition or processing system in an environment containing a variable level of ambient acoustic noise, which method comprises controlling the gain of the speech signal in accordance with an estimate of the speech level, said estimate being obtained by measuring said ambient acoustic noise level, and processing said measured acoustic noise level to produce an estimate of the likely speech level of said utterance.

1/1


SPEECH RECOGNISER
PROCESSOR
SPEECH LEVEL ESTIMATOR
20
22
18
24
26
16
DIGITAL GAIN
A-D
14
VOICE PROCESSING
12
10

Ref: DRA.3.USA

COMBINED DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **Speech Processing**

the specification of which: *(check one)*

REGULAR OR DESIGN APPLICATION

[] is attached hereto.

[ ] was filed on _____ as application Serial No. ____ _____ and was amended on _____ (if applicable).

PCT FILED APPLICATION ENTERING NATIONAL STAGE

[X] was described and claimed in International application No. PCT/GB99/03322
filed on 15 October 1999
and as amended on _____ (if any).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

PRIORITY CLAIM

I hereby claim foreign priority benefits under 35 USC 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

PRIOR FOREIGN APPLICATION(S)

Country	Application Number	Date of Filing (day, month, year)	Priority Claimed
Great Britain	9822529.5	16 October 1998	Yes

(Complete this part only if this is a continuing application.)

I hereby claim the benefit under 35 USC 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of 35 USC 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37 Code of Federal Regulations §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial No.) (Filing Date) (Status--patented, pending, abandoned)

POWER OF ATTORNEY

The undersigned hereby authorizes the U.S. attorney or agent named herein to accept and follow instructions from **WYNNE-JONES, LAINE & JAMES** as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the U.S. attorney or agent and the undersigned. In the event of a change in the persons from whom instructions may be taken, the U.S. attorney or agent named herein will be so notified by the undersigned.

As a named inventor, I hereby appoint the following attorney(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith : **Robert J. PATCH, Reg. No. 17,355, Andrew J. PATCH, Reg. No. 32,925, Robert F. HARGEST, Reg. No. 25,590, Benoît CÂSTEL, Reg. No. 35,041, Eric JENSEN, Reg. No. 37,855, and Thomas W. PERKINS, Reg. No. 33,027,** c/o YOUNG & THOMPSON, Second Floor, 745 South 23rd Street, Arlington, Virginia 22202. Address all telephone calls to Young & Thompson at 703/521-2297.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or first inventor: **Melvyn John HUNT**
(given name, family name)

Inventor's signature M. J. Hunt Date 6 April 2001

Residence: **Cheltenham, United Kingdom** GBX Citizenship: **British**

Post Office Address: **8 Dewey Close, Woodmancote, Cheltenham, Gloucestershire, GL52 4UF, United Kingdom**